ACTION EXPLANATION AND THE NATURE OF MIND*

1. INTRODUCTION

Action explanation has long been seen as a guide to psychological theory. The classic argument is Hume’s. Hume’s distinction between reason and the passions rests on the principle that reason is motivationally inert: “The reason alone can never be a motive to any action of the will; (...) it can never oppose passion in the direction of the will”. So to explain an action fully, Hume thinks, we must always appeal not only to an agent’s ‘reason’, or beliefs, but also to their ‘passions’, desires, or ‘pro-attitudes’.

Given that we want our psychological theory to be capable of explaining action, the principle that belief is motivationally inert thus puts a fundamental constraint on our view of the gross architecture of the mind. This principle is as orthodox today as it was controversial for Hume. As a result, despite its distance in other respects from Hume’s view of the mind, contemporary cognitive theory takes his ‘two-box’ psychology virtually for granted. (By way of corollary, Hume’s principle continues to be applied in areas of philosophy that depend on this psychology. In particular it is still taken, as Hume himself took it, to provide a foundation for various brands of ethical non-cognitivism. Briefly, the non-cognitivist wants to distinguish moral commitments from genuine beliefs; given Hume’s principle, it is sufficient to show that such commitments are active, rather than inert. More on this in section 5.)

I am going to argue that this contemporary reliance on Hume’s principle is unjustified. The most charitable contemporary reconstructions of Hume’s arguments fail to show that beliefs alone can never motivate intentional action. As a result they fail to show either that action explanation requires a two-box cognitive theory; or that moral commitments (even if agreed to motivate actions) are not beliefs.

However, my approach will be rather indirect. It seems to me that there is a similarity between Hume’s argument and some much more recent attempts to show that action explanation puts constraints on the nature of mind: arguments to the effect that action depends on demonstrative or indexical beliefs. The similarity seems to me to be of consid-
erable interest in its own right, as well as a good guide to the fundamental weaknesses of Hume's argument. So I shall first consider the demonstrative and indexical arguments, and try to show that they are invalid, at least as interpreted as constraints on arbitrary cognitive systems. I shall then argue that contemporary arguments for Hume's principle rely on similar moves, and hence fail for much the same reasons. The detour will be justified, I hope, by the interest of the scenery along the route.

2. ACTION AND INDEXICAL THOUGHT

In a paper entitled ‘Demonstrative Thought and Psychological Explanation’ Christopher Peacocke argues for what he calls the Indispensability Thesis:

IT: No set of attitudes gives a satisfactory psychological explanation of a person's acting on a given object unless the content of those attitudes includes a demonstrative mode of presentation of the object.

As it stands, this is a claim about persons, rather than about arbitrary cognitive systems. Even so, it admits of more than one interpretation: Is it a claim about what the minds of human agents actually do contain, or about what they could conceivably contain, given their organic limitations? If the latter, is there anything to prevent us generalizing to arbitrary cognitive systems — does the reference to ‘a person’ do any work?

The argument Peacocke actually offers for IT is not much help here. It consists largely in examples: in cases in which it seems intuitively clear that an action explanation does require a psychological attitude with demonstrative content. Such examples may convince us that certain human actions are in fact inexplicable without appeal to demonstrative thoughts: but it seems unlikely that they could show that the same would be true of the actions of any conceivable cognitive system, or even of any human agent. However, we shall see that Peacocke's examples do suggest arguments for a more general conclusion — arguments for regarding IT as a constraint on arbitrary cognitive systems.

A similar conclusion to Peacocke's, though concerning indexical rather than demonstrative beliefs, seems to have been widely drawn from John Perry's influential paper, 'The Problem of the Essential Indexical'. Perry argues that in order to explain why an action is performed when, where and by whom it is, we need in general to ascribe an indexical belief to the agent concerned. Like Peacocke, Perry argues by example, describing cases in which it seems clear that without indexical beliefs, we do not have an adequate explanation of a particular action. Again such examples show at best that certain human actions are in fact inexplicable without indexical beliefs, and not that the same would be true of the actions of any human agent, let alone of any conceivable cognitive system. This may be all that Perry himself requires. But again his examples suggest an argument for a more general conclusion (and have, I think, been taken to do so).

The most plausible and distinctive of these arguments runs something like this. An action is an intentional change in a person's (or cognitive system's) behaviour. To explain it we need to explain why the change happened when, where and in whom it did happen. This requires a change in the agent's cognitive apparatus (hereafter, ‘Cog’): the action is a result of something that happens in Cog, and uniformity there cannot explain a difference elsewhere. But in general it is only the agent's indexical belief states that do change in the required fashion. For example I may have believed for weeks that I was to start reading this paper at 11 o'clock on the third Tuesday in May. That belief did not change at the moment I started. What did change in me, and hence what explains my starting, was that I acquired the indexical belief that it was now 11 o'clock on the third Tuesday in May. That indexical belief thus played a crucial role in the production of my action. It filled an otherwise mysterious gap in the chain of psychological events, or states, that led me to open my mouth.

I shall call this the Causal Role Argument. With some quails, but for the sake of argument, I shall attribute it to Perry. Since the argument applies as much to possible robots as to actual humans, I thus take Perry to be arguing that indexical beliefs are essential in principle, as well as in practice.

Perry treats demonstrations as a kind of indexical. In one of his examples an agent has to believe 'that this is the Mt Tallac trail and that is Gilmore Lake'. So Peacocke's conclusion would seem a special case of Perry's. Whether the inclusion is proper will depend on what Peacocke means by an object (as of course will the scope of Peacocke's argument in general). However, if times, places and people are all to be
treated as objects, then the two conclusions might be expected to be co-extensional.

However, this ignores what might be thought a more significant difference between the two conclusions: that they seem to focus on different aspects of the explanation of action. Perry seems concerned with the causal role of an agent's awareness of a time, place, person or other object, in the production of an intentional action; Peacocke with the 'directedness' of an action towards a particular object. The Causal Role Argument seemed to rest on the intuition that an intentional action is an effect, *inter alia*, of certain features of an agent's environment. Since the immediate causal history of an action lies in Cog, relevant features of an agent's environment must be represented in Cog, in order to explain intentional behaviour. Perry's conclusion then rests on examples that suggest that only indexical belief states are appropriately sensitive to certain environmental features. Peacocke's conclusion, however, involves the notion of 'an agent's acting on a given object'. This suggests a concern with the effects of an action, rather than its causes. Perhaps more importantly, it seems to involve the idea that these effects may be *directed at* or *targeted on* particular objects. The problem is to cash this directedness in terms of the contents of Cog (and the claim is that only demonstrative thoughts will do the trick).

I think this difference will turn out to be fairly superficial. If an action is directed towards a given object in an agent's environment, that object presumably plays a causal role in the production of the action. If the object had not been there, the agent would not have acted in that way. Conversely, any feature of an agent's environment which, via a representation in Cog, plays a causal role in the production of an action, has a claim to be a feature towards which that action was 'directed'. Unless directedness amounts to more than an agent's intentional involvement of one object rather than another, we would seem justified in speaking of actions as being directed towards times and places, as well as to more concrete objects.

At present, however, there is nothing to be gained by insisting on assimilating the two conclusions. In the next section I shall therefore examine possible arguments for Peacocke's IT. The following section will then look at Perry's argument; and show, I think, that in so far as it rests on different grounds, it still does no better than Peacocke's.

3. IS 'THAT' REALLY NECESSARY?

Peacocke's main example involves an agent who reaches for a container that he believes to contain life-saving pills. We want to explain why the agent reached for that particular container. It seems we need to say that he believed that it was *that* container that contained the pills; or at least that it was *that* container which met some other description (such as 'left by the doctor') that he took to entail that it contained the pills.

The thought behind the Indispensability Thesis is that however many descriptive modes of presentation 'the Φ with which we supply the agent for his thoughts 'the Φ contains the pills I need', we will not supply attitudes sufficient to explain his acting on the container in front of him unless the agent has a thought with the same predicate 'contains the pills I need' in which that container is presented demonstratively.

The question here is: Why did the man reach for *that* object? A first thought might be that IT is somehow a product of the way in which we have framed the question - of the fact that in characterizing the action to be explained, we have ourselves referred to the object demonstratively. We can check this by substituting some other mode of presentation of the object concerned. Why, for example, did the agent reach for the bottle that used to belong to Uncle Fred? Again, because he believed that it - the bottle that belonged to Fred - contained the life-saving pills (and he wanted to live). Does this mean that object must have been presented to the agent as the bottle that belonged to Fred? Clearly not: in the last sentence but one the referring expression is used transparently. The agent believed of what we can refer to as 'the bottle that belonged to Fred' that it contained the pills; but he himself need not have referred to it this way.

Why then should we not say that in agreeing that the agent believed that *that* container the pills, we mean the demonstrative to be taken transparently? And hence that the agent need not have thought of the bottle demonstratively? The answer seems in part that we take for granted that the agent would have thought of it demonstratively - or at least would have had access to such a mode of presentation. He might not have known it as the bottle that used to belong to Fred, but he must have known of it as *that* bottle (or perhaps simply as *that thing*).

In other words, IT might seem to depend on what we might call the Availability Assumption:
AA: An agent who is acquainted with an object in such a way as to make it a possible target of an intentional action, is also sufficiently acquainted with it to refer to it demonstratively.

This principle has some plausibility, I think. An object sufficiently differentiated from its surroundings to be the target of an intentional action, seems sufficiently differentiated to allow demonstrative reference. However, it is not clear that AA guarantees IT. It is one thing to have access to a mode of presentation of an object; another for that mode of presentation to play an indispensable role in psychological explanation. It seems to me that there are two possible arguments for the extra step. I shall call them the Argument from Doubt and the Targeting Argument.

The Argument from Doubt relies on the intuition that (perhaps in virtue of AA) demonstratives provide a kind of standard, in terms of which it is held to be possible to establish the referents of other referring expressions. Thus in general demonstrative expressions enable us to enquire about – and more significantly, to doubt – the designations of other expressions. ‘Is that George (the last bus, the price of fish)’?, we can ask, when we don’t know what ‘George’ (‘the last bus’, ‘the price of fish’) refers to; or, ‘Is that really George (the last bus, the price of fish)?’, when we doubt it.

This doubt-enabling feature of demonstrative modes of presentation provides an argument for Peacocke’s thesis, along the following lines. The man who reached for the pills could have doubted whether that container was the one that contained the life-saving pills. In the grip of such doubt, he would not have reached for it (or at least his doing so would not be explained by the beliefs and desires we are presently invoking). So to explain his action we need to assume that he did not have such a doubt – in other words, it seems, that he believed that that container did contain the pills. In other words, he must have thought of the container under a demonstrative mode of presentation.

Note here the use of the assumption that a demonstrative mode of presentation of the container is available to the agent. We might try to argue (i) that if the agent had doubted that the bottle that used to belong to Uncle Fred contained the pills, he would not have acted as he did; and therefore (ii) that explanation of his action must invoke a thought in which the container is presented to him as the bottle that used to belong to Fred. However, this version of the argument trades on an equivocation between a transparent and an opaque use of the referring expression ‘the bottle that used to belong to Fred’. (i) is only plausible under a transparent reading; whereas (ii) requires that (i) be justified as an opaque attitude ascription. We have no reason to think that the agent knows the container under this description; and hence none for thinking that the belief he would express by ‘The bottle that used to belong to Fred contains the pills’ is a precondition of his action.

Thus in the demonstrative case AA saves the Argument from Doubt from what would otherwise be a fallacy of equivocation. The argument fails for other reasons, however. For one thing, it relies on the assumption that the lack of a disbelief entails possession of the corresponding belief – in other words that we believe whatever we do not doubt. In the case considered, this assumption is required to support the inference from the negative claim that the agent concerned did not doubt that that bottle contained the pills, to the positive conclusion that he believed that that bottle did contain the pills.

Do we believe everything that we do not doubt? In general, clearly not. Many propositions are undoubted in virtue simply of never having been entertained. To say that in such a case we believe whatever we do not doubt would be to ascribe contradictory beliefs to ourselves – for in such cases the negations of the propositions concerned have also not been entertained. The general principle needs to be qualified. We want to allow for unconscious beliefs – cases in which the lack of a pattern of behaviour that would reveal a doubt, is evidence of an implicit belief. But we don’t want to say that lack of any doubt whatsoever is equivalent to possession of a corresponding belief.

On any account, a crucial factor will presumably be the availability of circumstances in which a doubt could be manifest. Lack of manifestation of a doubt is to be taken to be manifestation of the corresponding belief. In this way the ascription of the belief will be given a behavioural justification. However, a lot will hang on the interpretation of the phrase ‘in which a doubt could be manifest’. For example, we might say that every time I eat an apple, I have the opportunity to manifest a doubt that it has not been poisoned. But it would be absurd to conclude that on every such occasion I believe that my apple has not been poisoned. If this were the principle of belief ascription that underlay IT, then satisfactory psychological explanation of action would be impossible. For there would be no end to the number of such beliefs we would have to ascribe, to explain any action whatsoever.
This points to a second and more basic flaw in the Argument from Doubt. Not only is the lack of a doubt not equivalent to a belief; but the explanation does not have to mention that lack — on pain, clearly, of having to mention the lack of endless doubts. The principle we need to preserve is this: an action \( A \) may be explained by attitudes \( P_1, P_2, \ldots, P_n \), even though it would not be explained by — indeed, would not have been performed given — attitudes \( P_1, P_2, \ldots, P_n \), and \( R \). There should be nothing surprising about this principle. It applies to causal explanation in general. Causes are usually at best sufficient in the circumstances for their effects. And with this principle in place, the Argument from Doubt is inadequate to support it.

It is unlikely that Peacocke intends it to rest on the Argument from Doubt. Despite its deficiencies, however, the argument does suggest a more subtle argument for it — for the two arguments share a common assumption.

Consider again Peacocke’s hypochondriac. Suppose now that like us, he believes that the pill bottle in question used to belong to Uncle Fred. ‘Why did you reach for the bottle that used to belong to Fred?’, we ask him. ‘I reached for the bottle that contained the pills’, he answers; ‘it happened to be the one that belonged to Fred’. In this case it is clear that although in fact he believed he was reaching for the bottle that used to belong to Fred, this belief plays no role in explaining his action. If he had doubted the same thing, it would not have made a difference.

If analogous reasoning applied to the demonstrative case, it would defeat the Argument from Doubt. Why is it then that when we ask the hypochondriac why he reached for that bottle, we should not be satisfied with the explanation: ‘I reached for the bottle containing the pills; it happened to be that one’?

The answer seems to be something like this: in order to direct an action towards a particular object, we rely on our perceptual access to the object in question. Targeted action requires perceptual guidance: an action cannot be directed at an object, unless that object is distinguished from other objects (and from any ‘background’) in an agent’s perceptual field. Moreover, if Cog is to control a targeted action, the target object must not only be represented in Cog; its representation must be suitably related to the object’s distinguishing perceptual characteristics.

The point can be illustrated by a series of simple examples. At one extreme, consider a hunter who sets passive traps for some sort of prey: a web-building spider, for example. When a spider catches a fly in this way, we would not say that the spider’s action was directed at that particular fly. On the contrary, the fly just happened to satisfy the various general parameters — size, spatiotemporal location, and so on — in virtue of which any object would be captured by the web. As in the case of a pond that is hit by a meteor, to use Peacocke’s example, the fact that it is that fly is a kind of accident of history.

Next consider an intermediate case. Like spiders, many frogs make a living catching flies. But they usually do it with their tongues, a feat that presumably requires perceptual control. The frog’s perceptual apparatus guides its tongue to a passing fly. So there is a sense in which this action is targeted on a particular object. The difference shows up in counterfactuals: if the fly had been in a slightly different place, the frog would still have caught it; but if the spider’s fly had flown a little to the left, it would have missed the web. (Of course there are limits: if the fly had been in a different swamp altogether, even the frog would not have caught it.)

However, the frog’s action is still not cognitively targeted. This is not primarily because the frog lacks a cognitive unit. Even if frogs were as intelligent as we are, their cognitive control of their tongues (qua fly-catchers, at any rate) could be expected to be limited to stop/start instructions. Feeling peckish, the frog’s Cog would thus issue the instruction: Catch a fly! Not ‘Catch that fly!’ for the perceptual system that guides the tongue to the fly just isn’t linked to Cog. Perhaps it used to be, but those frogs weren’t so good at catching flies — by the time they had finished cogitating, the fly had usually flown. So the intentional act of fly acquisition is better made non-specific — any fly will do. This means that although the bodily movement is perceptually targeted to a particular fly, from Cog’s point of view it is still an accident of history that a particular fly turns up for lunch.

Full cognitive targeting thus seems to require continuous cognitive control of a perceptual guidance unit. The crucial point — the crux of the Targeting Argument — is that in such a case the cognitive mode of presentation of the targeted object must be appropriately linked to the perceptual distinctions employed by the perceptually-guided targeting unit. Cog can be said to control a targeted action only if several conditions are satisfied. For a start, Cog must be capable of discriminating the object concerned from its background to the same extent as the relevant perceptual apparatus (i.e., as that part of the agent’s perceptual
apparatus that is linked to the relevant motor systems). But mere discrimination isn't enough – a frog Cog might be capable of following the fly-catching operation, without controlling it. Cognitive control requires a suitable flow of information between Cog and the perceptually-guided targeting unit. It would seem that Cog must be capable of representing the target object in perceptually-grounded terms. Cog must make a selection from a range of alternatives presented by perception – in effect, it must point to something in its perceptual field. Otherwise it will once again be as if Cog had issued the general instruction: Find something – anything – that matches the following description: . . .

At this point demonstrative modes of presentation seem just the job. Demonstrative reference seems perceptually-grounded. Demonstratives seem to work by calling attention to distinctive features of a person's perceptual field. In principle any perceptual distinction seems a possible basis for such a mode of presentation – and in particular, any distinction exploited by a perceptual guidance system. Demonstratives thus seem to supply the right kind of link between Cog and perception. And on this basis the Targetting Argument claims that unless an object is picked out for an agent by a demonstrative mode of presentation, then an action cannot be considered to be intentionally directed at that particular object.

However, it is one thing to identify a function that demonstratives are well-suited to serve; another to show that only demonstratives will do the job. I am going to argue that the required action-guiding link can be (and sometimes is) established in other ways; and hence that IT does fall in certain (admittedly exceptional) cases, and that it could fall much more widely. I begin with an example which, though perhaps not a counter-example (it depends on what we mean by an object), at least suggests where we should look for counter-examples. The example involves the action of choosing one's words, of deciding to express oneself in this way rather than that.

_G_ describes _F_ as 'Kerr's cur'. An unusual phrase – why did _G_ choose it? We can imagine that before his mind's eye, as he stepped forward to speak, _G_ could see a number of fitting ways of describing _F_. Why did he choose that one? Clearly he wished to describe _F_ in derogatory terms. But he would be speaking in public, indeed to history. The occasion called for a memorable phrase. Given _F_’s association with a man named Kerr, 'Kerr's cur' must have seemed (as _G_ put it, years later) 'very apt'.

Do we need to add that _G_ believed that that phrase was the phrase

"Kerr's cur"? Apparently not, so long as we avoid a type-token confusion. What needs to be explained is _G_’s choice of an expression of a certain type (from the range of epithets that occurred to him as he reached for the microphone). His use of a particular token is in general fully explained by his choice of the type. True, there could be cases in which we would need to invoke _G_’s belief that _that_ phrase – the noise that came from his lips – was a token of a certain phrase-type. But these are irrelevant to the present case, in which what we want to explain is the choice of the phrase-type in the first place.

What happens as _G_ makes his choice? Let us imagine it is something like this: he mentally surveys a shortlist of candidates (already selected in some other way). Having made his choice, he loads the chosen phrase into a appropriate speech register. The question is: In loading the chosen phrase, does _G_ have to think of it as 'that phrase'? It seems not – and the reason seems to lie in the similarity between on the one hand the phrases themselves, as they appear on _G_’s mental list, and on the other their modes of presentation in Cog. For it is this connection that would make it absurd to object that _G_ chose, as it were, only the description of a phrase, and that the actual phrase he used just happened to be the one that matched this description. The relationship between words and their names is no mere accident.

It might seem that the example hangs on the fact that all the action takes place in _G_’s head – and that it is only in a metaphorical sense that _G_ can be said to 'choose' or 'pick out' the phrase 'Kerr's cur'. But suppose that _G_ actually had a written list of suitable phrases (his speech writer having thoughtfully jotted down a few suggestions). Before he speaks, _G_ runs through the list, and draws a circle around 'Kerr's cur', to remind himself to use it. His circling the phrase is presumably a perceptually-guided action. What instruction did he issue to his perceptual guidance system? Was it 'Circle that one', or simply 'Circle "Kerr's cur"'? The important thing is that the latter would do. In effect, it presents the perceptual guidance system with something that it can match against its own perceptual information – a picture of the targetted object. (The instruction is: Find something that matches 'Kerr's cur', and then circle it.) So there is no longer an essential role for the demonstrative thought. Of course, _G_ could issue a demonstrative instruction, but he doesn't have to.

All the same, it may seem that no objects more concrete than words can be targetted non-demonstratively: and hence that IT holds, so long
as suitably restricted to 'proper' physical objects. These days we are well
aware that there are few interesting similarities between names and the
objects they refer to. Even onomatopoeia is rather uncommon. Names
are normally the products of quite arbitrary linguistic conventions. In
other words, the representations of objects in Cog are likely to bear no
intrinsic relation to the perceptual images of the objects they
represent. Only in rare cases, such as when the referent is itself a
word, are linguistic representations anything more than arbitrary
symbols. Language is generally non-perceptual (or should we say that
perception is generally non-linguistic?).

However, we should avoid the fallacy of unwarranted generalization.
After all, we are more interested in principled constraints on arbitrary
systems than in contingent human limitations. The problem,
we have seen, is to secure an appropriate action-guiding link between
cognition and perception — roughly speaking, to provide a flow of
information between the two, in a form usable in cognition, and of such
a content as to allow perceptual guidance of an action towards a
particular object. And the difficulty seems to stem from the fact that
there is generally no natural similarity between the representation
of information in perception and its more symbolic representation in
cognition. So it is as if we are required to link two different transport systems:
the link has to match one to the requirements of the other. Let us
consider the range of possible solutions to such a problem.

The transport analogy suggests two basic solutions: from a given base,
either system can be extended to meet the other. Cognition can be
extended to employ as linguistic symbols, or forms of representation,
the existing outputs from perception. Or perception can be extended to
supply outputs in a form already appropriate to the symbolic capabilities
of cognition. We shall see that both strategies are capable of providing
non-demonstrative targetting of objects of many kinds. In principle
these strategies could be combined, producing a wide range of hybrid
solutions. In transport terms this would correspond to extending both
systems in order that they should meet in the middle.

There is a third basic strategy, that I shall not explore in any detail.
It corresponds to the construction of a bridging system, matching the
output of one transport network to the input of the other. At least one
implementation of this strategy would exploit the role of 'similarity' in
connecting cognitive and perceptual symbolism. The strategies we are
going to discuss will turn on modifications to cognition, perception, or
both, which would have the effect of producing such similarities. But the
notion of innate or objective similarity is highly problematic. Similarity
often lies in the eye of the beholder. This suggests that the required links
might be achieved by modifications in our 'similarity sensors'. In other
words, the required associations between cognitive and perceptual
representations might simply be innate. We might be built to perceive a
'natural' connection between redness and the word 'red'. In what
follows, however, I shall rely on our existing intuitions about the
similarities between modes of representation.

The first strategy involves the use of symbols that are not conven-
tional words, and whose meanings depend on (what we take to be) their
natural similarities to our perceptual images of the things they refer to.
This sort of thing is actually quite common. In speech, we often refer
to a sound simply by mimicking it. And in writing, we quite often intro-
duce visual representations, or likenesses, of something to which we
want to refer. Here are a couple of examples, one written and one
meant to be spoken:

(i) The child grabbed the steering wheel and said [brooom, brooom]
   (here I imitate a child's imitation of an engine noise).

(ii) Someone has painted [X] on the wall of our embassy.

It is not difficult to see that as in the 'car' case, explanation of the
actions described in these examples might well have no need of a
demonstrative mode of presentation of the noise or the peace symbol.
For example, we can say that the child wanted to make a noise like a
car, and believed that [brooom, brooom] is a such a noise. It would be
pointless to add that she believed that that noise (i.e. the noise that she
made) was the noise [brooom, brooom].

Clearly the number of such cases will depend on what sorts of things
can be used as cognitive symbols. A language based on pictorial images,
for example, would enable an agent to target ordinary physical objects.
In a way this is what demonstratives provide. They direct us to some-
thing in our perceptual representation of the context of utterance,
enabling that context to play a role in determining the meaning of what
is said. The advantage of such a system lies in its extraordinary gener-
ality, or universality. It enables us to generate a symbol for virtually
anything in our physical environments, provided only that it is distinc-
tive enough to stand out in perception to ourselves and our fellow
speakers. And it thus provides linguistic access to an environment whose diversity would otherwise far outstrip the symbolic resources of human language.

However, demonstrative reference achieves its universality at some cost. In communication it depends on the existence of a shared perceptual framework. And it has the usual disadvantage of a multi-purpose tool: to all intents and purposes, it can only do one thing at once. It is quite conceivable that particular cognitive systems, in particular circumstances, would be better served by more specialized linguistic tools: for example, by the ability to use pictorial images themselves as linguistic symbols. This would differ from the demonstrative case in allowing the image itself to function as a symbol, rather than employing a surrogate symbol, or 'pointer', whose function is to call attention to something in the perceptual field.

Consider for example the cruise missile, which is said to employ pattern-matching techniques in locating its target. Its basic instruction takes the form 'Go to [picture]'. The missile is designed to fly to something that looks like [picture] (and then explode). It is true (one hopes) that existing cruise missiles are no more than perceptually-guided motor units - their instructions come from outside. But more advanced models might choose their own targets, and give themselves instructions of the form 'Go to [picture]'.

So much for extending the symbolic repertoire of the cognitive system. The second approach is to enhance the perceptual apparatus, in order to make perceptual information available in Cog's existing symbolic format. An example in the same vein as the last: bomber pilots used to depend on demonstrative presentations of their targets. 'Das ist London', the Luftwaffe presumably thought, as they steered towards the glow in the western sky. Modern technology has made such thoughts redundant. Modern military aircraft carry navigational computers, linked to a device to project a three-dimensional image in the pilot's visual field. The computers calculate the aircraft's position in relation to its target, and might for example project two small circles, which appear to the pilot superimposed on the visible landscape. One circle appears over the target, and is labelled 'TARGET'. The other appears at infinity in the aircraft's present direction, and is labelled 'PRESENT DESTINATION'. To reach the target, the pilot need only steer the aircraft so as to keep the two circles concentric. The pilot is still a human agent, capable of directing intentional action towards objects in the external world. The difference is simply that thanks to the enhanced perceptual system supplied by the aircraft, it is not. Some of the pilot's perceptually-guided intentional actions do not depend on demonstrative presentation of his or her environment.

Suppose for example that the target is a pond, and the pilot successfully bombs it, having thought of it simply as 'TARGET' (the description used by the enhanced perceptual system provided by the aircraft). Is the fact that that particular object was bombed a kind of accident, reflecting the fact that it 'just happened' to be the object that matched that description? No, for the description is perceptually linked. Even if the pond had been somewhere else the description would still have applied to it - which means that in virtue of the fact that the corresponding cognitive symbol would have had a different reference, the pond would still have been bombed.

At this point it may be objected that if the computer had picked out something else as the target, the pilot would have directed both action and aircraft towards some other object. Doesn't this mean that the fact that the pilot bombed a particular object is not explained by the pilot's cognitive state - but only by the conjunction of this and the fact that the computer picked out the object in question? I shall return to this objection in a moment. (In fact there is nothing here not already a product of ordinary perceptual systems.)

The bomber example trades on the fact that in their professional capacity, at any rate, bomber pilots are called on to perform a very limited range of actions, on a rather limited range of objects. In other words, theirs is just the sort of environment in which demonstratives might be expected to be dispensable. For we saw that demonstratives are multi-purpose names - in effect, a single universal name - and the value of a multi-purpose tool is proportional to the number of tasks at hand. In an effectively finite and perceptually well-discriminated environment, a cognitive system may well do better with individual perceptually-grounded names.

One more example, to illustrate this point: consider the central taxation computer. It inhabits a universe of taxpayers, perhaps divided into a number of natural kinds: individuals, corporations, charitable bodies, and so on. The computer knows each entity in this universe of taxpayers by its own unique name. All the information the computer
receives about its universe is expressed in terms of these names; as are all the instructions it issues concerning particular taxpayers. There is simply no place for demonstrative thoughts.

Does this mean that the computer is incapable of directed action, of action targeted on a particular taxpayer? In favour of this view, it may be said that to explain the fact that the computer acts on a particular individual - issues me with a demand for a billion dollars in back taxes, for example - we need to add to our account of the computer's cognitive state, at least the information that I am what the tax office refers to as 512/879/3274. The computer merely instructs its subordinates to send 512/879/3274 a demand for a billion dollars. To explain the effect of this instruction, we then have to look outside the computer, and tell the causal story in virtue of which the letter in question ends up in my mailbox.

This is the objection we deferred in the previous example. There, the corresponding claim was that to explain why the pilot bombed a particular object, we have to note that the aircraft's computer picked that object out as the target, rather than something else. The objection threatens the entire strategy of perceptual enhancement. An enhanced perceptual system, it suggests, would leave one curiously out of touch with the world - so much so that to explain one's actions, we would need to refer to the mediating perceptual system. Indeed, it is easy to see which slippery slope the objector thinks we are on: such a perceptual system might simply be deceiving the cognitive system that possessed it. The pilot might simply be operating a rather realistic flight simulator - while the unfortunate tax computer might simply be sitting on a test bench, the electronic equivalent of a brain in a vat.

In fact, however, this slope slopes the other way. The appeal to the possibility of illusion backfires, because in so far as it is problem at all, it affects ordinary perceptually-grounded explanations as much as these imaginary cases. Suppose for example that among his other problems, Peacocke's hypochondriac suffers from hallucinations. He sees bottles of pills the way desert travellers see cases. When at last he grasps the real bottle, we need to add to our explanation a note to the effect that in this case there was a bottle where he thought he saw one. (Whether we regarded this as a kind of accident, or whether we said that he had actually perceived the bottle, would presumably depend on the details of his condition.)

Ordinarily we take for granted that other people's perceptual systems function reliably, and much as our own. Without this tacit assumption, demonstrative communication would be impossible. With it, we have the Availability Assumption, and a part of the background of ordinary action explanations, which determines what we take to call for explanation. In effect, it means that action explanations are not usually required to unpack demonstrative references. But it does not mean that there is nothing to unpack, no causal story to be told about the role of demonstrative thought, perception and the world in the production of action. There is a story, as there is in our perceptual enhancement examples. Whether an explanation has to tell it depends in every case on what were, prepared, or allowed, to take for granted. Demonstrative thoughts have no favoured status here - or at any rate none sufficient for Peacocke's Indispensability Thesis. Where perception does not discriminate between objects, not even demonstrative thoughts can explain the fact that we act on one rather than the other.

The Targetting Argument thus offers a promising account of the special role that demonstrative thoughts seem to play in action explanation. But at the same time it leads us to a variety of actual and possible counterexamples to the Indispensability Thesis. It seems to be false in all its various interpretations, and certainly so in its strongest and most interesting ones. In the general case, directed intentional action does not seem to depend on demonstrative thought. Demonstratives merely provide one of a number of conceivable solutions the problem of cognitive representation of the possible objects of intentional action. The problem has several variables: the size and variability of the universe of action, the symbolic resources of cognition, and the nature of the perceptual system. In general the advantages of demonstrative modes of presentation will depend on the balance of these factors. In particular cases - perhaps in our own, most of the time - they may indeed provide the only solution. But if and when this is true, it is clearly due to the peculiarities of the case in question, and not to any general need for demonstrative thoughts.

4. I, ROBOT?

I think that a similar strategy can be used against the Causal Role Argument, which in section 2 I attributed (somewhat tentatively) to Perry. That argument relied on two main claims: firstly that if actions are to be caused by the contents of $C_{og}$, then actions must be correlated
with changes in Cog (in response, in turn, to changes in relevant environmental factors); and secondly that only indexical belief states show the appropriate sensitivity to environment – non-indexical beliefs being too coarse-grained to explain many of the differences between the intentional behaviour of different agents, and of the same agent at different times.

Perry's most striking examples involve sensitivity to temporal changes, and to questions of an agent's own identity. In both cases it may be difficult to see how the sort of speculative counterexamples we used against the Targetting Argument are going to apply. Let us begin with times.

We described two ways in which non-demonstrative modes of presentation might target environmental features. One involved an extended symbolic language, capable of utilizing the outputs of perception without the need for demonstratively-grounded bridging conventions. The other involved an extended perceptual apparatus, capable of supplying a symbolic output of a kind presently utilized by Cog.

In a world in which times are as poorly distinguished as they are in ours, it is indeed difficult to see how times could be 'targetted' by a modification of the former kind. We do not appear to perceive times, as distinctive features of our perceptual fields. (Partly for this reason, it seems a mistake to try to treat temporal indexicals as demonstratives to paraphrase 'now' as 'this time'. Temporal indexicals seem rather to rely on the fact that our perceptual experience - indeed our mental life as a whole - has a temporal parameter.) However, it is easy to imagine an environment that would allow cognitive systems to perceive times, in quite a natural sense: all that is necessary is that the environment contain some monotonically changing property, to which the cognitive systems concerned are perceptually sensitive. In a community of short-lived but sharp-eared creatures, a slowly changing musical tone might do the trick. So long as these creatures used mental representations of musical tones as cognitive symbols, it seems to me that they would not need temporal indexicals.

In any case, there are much more natural examples of the latter sort of modification, which consists in our perceptual apparatus being modified to suit our existing cognitive symbolism. Times have non-indexical names, provided by the standard calendar. Perry's examples show that for us, calendar dates and clock readings do not target the moments they refer to. Non-indexical temporal beliefs do not explain the fact that we act at one time rather than another. Conversely, we can target a time - intend to act at a particular moment - without knowing what calendar time it is. Even in the age of the digital watch, it would seem that most deliberately timed human actions are performed in ignorance of the information that such devices provide. This is hardly surprising, since our cognitive apparatus has presumably evolved in environments in which (the Great Watchmaker notwithstanding) useful timepieces were exceedingly rare. But the contingent constraints of past environments should not be confused for theoretical limitations. Whether robots or humans, we might in future do things differently. In particular, we might have accurate internal clocks, whose output would be constantly accessible to cognition. As in the pilot case, for example, this clock might be linked to an overlay on the visual field, so that at any time we would have direct visual access to a non-indexical symbol by means of which to refer to that time.

Imagine that we are built like this, and that we want to explain why I started reading at 11 o'clock. We mention my belief that the paper was scheduled to start at 11 o'clock, and my desire for punctuality. Is it necessary to add that at 11 o'clock I came to believe that it was then (i.e. 'now') 11 o'clock?

It seems to me not. However, there is an appealing argument to the contrary, stemming from the causal notion of explanation with which we are here concerned. (Indeed, it is an elaboration of the Causal Role Argument.) The function of indexical belief states is supposed to be to make the contents of Cog sensitive to a certain sort of environmental change: in the temporal case, either to changes in environmental indicators of time, or to 'changes in time itself. (Here we can safely ignore the metaphysical problems associated with the notion of the passage of time.) The model is roughly this: an environmental change-in this case, a change in the time) is perceived by the agent, and registered as a change in an indexical belief state. (I come to believe that it is now 11 o'clock.) The newly-acquired indexical belief state then interacts with various non-indexical attitudes, and hence leads to an intentional action.

Thus the non-indexical part of Cog is being thought of as dispositional. It is disposed to respond in certain ways to changes in the indexical sector of Cog. The important question is whether what triggers the disposition itself to be part of Cog. If so - and the Causal Role Argument depends on the claim that it is so - then Cog must contain the sort of state that is capable of exhibiting this kind of environmental
sensitivity (and an overlay clock would not eliminate the need for temporal indexicals).

The principle that the triggers of action lie in Cog might be thought to follow from two independently plausible intuitions: first, that changes in effects require changes in causes; and secondly, that the environmental causes of intentional actions are mediated by the contents of Cog. Taken together, these intuitions may seem to require that relevant environmental changes must themselves be represented in Cog.

However, this argument is self-defeating; for there is an analogous argument involving the indexical and non-indexical parts of Cog. Changes in the indexical part (if any) of Cog are among the causes of intentional action. And their effects are mediated by the non-indexical part of Cog. But clearly we do not want to say that changes in the indexical part must be reflected in changes in the non-indexical part. The very purpose of the indexical part was to quarantine these changes.

To avoid this reductio of the indexical/non-indexical division of Cog, we presumably need to say something about the mediating role of the non-indexical part: for example that it acts like a conduit, conveying changes from the indexical part of Cog to the Will without itself changing in the process. Something like this will presumably make sense of the two intuitions we started with. But now the problem is obvious: why shouldn't we apply the conduit model in the first place, rather than introducing the indexical part of Cog? Why shouldn't we think of Cog as a whole as a conduit, conveying changes in our perceptual states to changes in the Will? The way this seems clear for cases in which the non-indexical part of Cog responds directly to inputs from perception—inputs that are not themselves cognitive attitudes.

The overlay clock provides one such case. Equipped with such a clock, we would apparently have no use for perceptual reports such as 'It is now 11 o'clock'. This is not simply the pragmatic point, that it won't occur to any of us to doubt that we would have this belief at 11 o'clock (so that from a pragmatic point of view, the explanation of my starting to speak at 11 o'clock would not need to mention it). It is rather the argument that there is no longer a causal role for any such belief. The non-indexical attitudes I mentioned earlier (the belief that the paper was due to start at 11 o'clock, and the desire for punctuality) together constitute a disposition to respond to a certain change in my visual field; that change that takes place when my overlay clock flips over from 10:59 to 11:00.

It might be objected that clocks are never totally reliable, and that even if one of them was reliable, it would always be possible to doubt its reliability: to ask, when the clock shows 11:00, 'Is it really 11 o'clock'? And doesn't this question depend on the possibility of demonstrative reference to times? In other words, isn't what is being doubted here indexical in form? By itself this objection is not an argument for the indispensability of indexical belief states in action explanation. But once the possibility of doubt is admitted, it soon becomes one. For we then feel inclined to say that when I started reading at 11 o'clock, I must have trusted my internal clock—consciously or otherwise. I must have believed that it was then 11 o'clock.

However, this is just the Argument from Doubt. It fails here as it failed as an argument for Peacocke's IT. First, it depends on the assumption that belief is entailed by a lack of the corresponding disbelief. Secondly, and more importantly, it tacitly relies on the assumption that only indexical or demonstrative reference can make the contents of Cog causally sensitive to features of an agent's environment. But of course the example is intended precisely to illustrate how that assumption might fail, for a suitably equipped community of agents. (True, the example requires that such agents take for granted the reliability of their perceptual equipment. But in the previous section we observed that any perceptually-grounded system of reference requires something like this. In our case it is the tacit assumption that our perceptual apparatus is sufficiently reliable, and shared, for demonstrative or indexical reference to work at all.)

So much for times. However, Perry's most striking examples concern an agent's awareness of his own identity. Perry imagines himself in a supermarket, spilling sugar from a broken packet. Seeing a reflection in a mirror, he may believe that that man is making a mess. He may even believe that John Perry is making a mess. But it is only when he comes to believe that he himself is making a mess (i.e. what he would express by 'I am making a mess') that he is moved to do something about the spilling sugar.

Dramatic as these sorts of examples are, it seems to me that there are no problems here not already encountered in the temporal case. If anything, personal indexicals are easier to eliminate than temporal ones; for self-identities, unlike times, are cognitive constants. Barring forgetfulness, knowledge of one's own identity need be gained only once. It follows that in principle perception need have nothing to do
with the acquisition of this sort of knowledge: we could simply be made with it. In other words, each of us could be pre-programmed to use a personal name, unique to ourselves, as we now learn to use the indexical 'I'.

Others could use our persona' names, for example to convey information about us to ourselves and third parties. Huw Price could thus inform John Perry that John Perry is making a mess. Both of us would then believe what we would both express by saying ‘John Perry is making a mess’. The rest of our relevant beliefs and desires may also match, but Perry acts and Price does not. How can this be, if both are cognitive systems, acting on the basis of their beliefs and desires? In virtue of the fact that in one crucial respect, the same cognitive systems are wired into different agents. Perry is wired to act on the basis of beliefs of the form ‘Perry should do X’ (or desires of the form ‘Perry wants to do X’), whereas Price is programmed to act on the analogous beliefs or desires about Price.

It might be objected that the example actually supports Perry’s case, in that it illustrates that non-indexical beliefs are not in general sufficient to explain action (since in such a case we have to refer to the sort of agent in which the beliefs and desires are embodied). However, I agree that non-indexical beliefs do not suffice to explain action. I am only denying that this entails that intentional agents must have indexical beliefs. The point of this example, as of the others, is simply to show that the explanatory work normally done by indexical (or demonstrative) beliefs can be done in other ways. Suitable-designed cognitive systems would not need such ‘locating beliefs’ (as Perry calls them). Instead, the task of ‘locating’ their entirely non-indexical Cogs at a particular point in their environment would be handled by their perceptual and motor components.

I conclude that so far as Perry’s argument differs from Peacocke’s, it is no more successful. In particular, the Causal Role Argument is invalid. Equipped with appropriate ancillaries, an agent can be sensitive to environmental changes, without representing in Cog the environmental differences which are relevant to action explanation. Indexicals, like demonstratives, are essential only for creatures of our present and thoroughly contingent limitations – and even then not all the time. As for robots, they would be well advised to help themselves to in-built clocks and pre-programmed personal names (and to on-board position sensors, for that matter); and hence – among themselves, at any rate – to escape indexical complexities.14

5. THE END OF DESIRE?

What does Hume’s argument have in common with those of Peacocke and Perry? Obviously all three appeal to action explanation, but is there any more to it than that? In this final section, I shall try to show that there is. Hume relies on the principle that beliefs are motivationally inert. I shall argue that this principle’s plausibility derives from arguments and assumptions that run parallel to some of those we have seen to underlie the arguments of Peacocke and Perry (and hence that it is vulnerable to similar objections).

First an objection: Hume’s principle might seem to be entailed by a perfectly orthodox view of the explanation of action. Almost everybody agrees that at least one way to explain an action is to describe the agent’s reasons for doing it – the beliefs and desires that led him or her to act in that particular way. In challenging Hume’s argument, wouldn’t we be challenging this whole picture of action explanation?

To the extent that this view of action explanation derives from the Humean model, I think it is true that in criticizing Hume’s principle we shall be offering reasons for a critical reassessment of the orthodox model. However, for present purposes we need not argue for the rejection of this model. For in one crucial respect Hume’s conclusion is stronger than the model requires. It would be quite in keeping with the orthodox view to say that the explanation of action involves two sorts of belief: beliefs about the desirability, or value, of certain states of affairs; and beliefs about the bearing of the action in question on the obtaining of those states of affairs. Hume’s principle, on the other hand, claims that belief alone cannot motivate action – in other words, that the explanation of action requires mention of states of mind that are not beliefs.

This difference is crucial to the two applications I mentioned in section 1. On the one hand, Hume’s principle is of interest to cognitive theorists precisely because it demands a category of cognitive states distinct from beliefs. If true, it is thus a major constraint on the large scale structure of cognitive systems. It implies that agents must embody at least two different ‘boxes’ of cognitive attitudes, and hence that as
theorists we need at least two distinct accounts of the relation of the content of an attitude to its cognitive role. If desires could be regarded as a kind of belief, things would seem much simpler. The task of explaining the distinctive cognitive role of a desire would then fall to the general theory of content (which has lots of explaining to do anyway), and we could hope for a uninary account of the relation of content to cognitive role.

Similarly, Hume's distinction between motivating desires and non-motivating beliefs is crucial to the use that is made of Hume's argument by non-cognitivists in ethics and aesthetics. Non-cognitivists want to show for example that moral judgements are not factual judgements in the ordinary sense. This requires a general distinction between fact stating and non-fact stating uses of language. Typically, at any rate, non-cognitivists rest this semantic distinction on a psychological one: genuine statements of fact are taken to be the expressions of genuine beliefs.

The non-cognitivists then want to show that moral statements lie on the right side of this psychologically grounded distinction — in other words, that such statements do not express beliefs. Given Hume's principle, it is enough to show that moral commitments do motivate actions; for this puts them on the desire side of Hume's distinction. And to many, at any rate, it "seems to be a conceptual truth that to regard something as good is to feel a pull towards promoting or choosing it." Clearly, someone who rejected Hume's principle could quite well endorse the standard model of action explanation, and accept that moral judgements express mental states on the active or motivating side, and yet insist that moral commitments are genuinely beliefs about moral facts.

Note, however, that such a view admits two possible accounts of the place of desires. One account would recognize a category of desires, distinct from beliefs; but claim that motivation is not the exclusive preserve of attitudes of this category (moral commitments being one sort of exception). The problem will then be to say what distinguishes beliefs from desires, given that it is not the power to motivate action. This problem might incline us to the second sort of account, which would recognize no useful distinction between beliefs and desires, claiming that all propositional attitudes can be represented as species of belief.

These two accounts agree in rejecting Hume's principle — in other words, in accepting the possibility of motivating beliefs. Why, these days (freed of the other peculiarities of Hume's philosophy of mind), should anyone deny this possibility? The first line of argument, it seems to me, is an analogue of the Argument from Doubt. We might call it the Argument from Dislike: for any property of an action or state of affairs, it is conceivable that an agent could be motivated by a desire that not F. This being so (the argument runs) no action can be explained simply in terms of a belief that the action concerned would be F or would produce some F state of affairs. We need to add that the agent desired that F, or at least did not desire that not F. In other words we need to mention the agent's desires.

I think it should be clear that this argument depends on the same kind of assumption as the Argument from Doubt: that the fact that an explanation would be inadequate in the presence of some extra condition, entails that it is inadequate unless it mentions the absence of that further condition. The assumption is no more plausible here than it was for demonstrative beliefs (or for causal explanations in general). As in the demonstrative case, to admit the assumption here would be to abandon any systematic approach to the contents of Cog. For there is no end to the things that even the most fastidious agent does not dislike.

From the point of view of the second of the two accounts we recently distinguished, there is another objection to the Argument from Dislike. For this account sees the question at issue not as whether agents have desires, but as whether these desires are anything but a special class of beliefs. On the latter view — the view the argument is intended to refute — to believe that an outcome of a possible action is F might be the very same thing as desiring that outcome. In this case the possibility that the argument imagines — that of someone holding the belief, but desiring an outcome such that not F — would be that of a person who desires what he takes to be undesirable. We seem entitled to regard this as no possibility at all.

At the corresponding point in our discussion of Peacocke's argument, we asked why doubt (in demonstrative form) should matter. If an agent wants the bottle containing the pills, why should it matter if he doubts whether that bottle contains the pills, more than it matters if he doubts whether the bottle that belonged to Uncle Fred contains the pills? The answer seemed to lie in the assumption that in virtue of its special
association with perception, the demonstrative thought plays a crucial role in enabling an action to connect with its target.

Is there an analogous question in the present case? I think that there is. The argument again turns on the connection between two mental components of a fully-fledged intentional agent: on the one hand Cog, the seat of cognitive deliberation; on the other, a unit whose function is to initiate bodily action, in response to the deliberations of Cog. There are a number of possible models of this latter unit — let us call it ‘the Will’ — and of its relation to Cog. At one extreme we could regard it as a kind of store, or processing house, for intentions already formed in Cog. At the other extreme we could take it to be the factory in which judgements arriving from Cog are matched with a store of desires to yield intentions. However, since we want a model in which the action which interests us takes place at the boundary between Cog and the Will, let us avoid both extremes. Let us view the Will as the surface on which intentions ‘condense’, as a consequence of reactions that take place in Cog.

In effect, Hume’s claim is then that no intention can condense in the absence of a desire. The Argument from Dislike claims that whatever beliefs are present, we can imagine a desire whose presence would prevent an intention condensing from those beliefs. Its rebuttal points out, first, that even if this were true, it might well be irrelevant to the question at issue, which is whether in the absence of that desire (or any other), these beliefs are causally sufficient for the formation of the given intention; and secondly that if we are allowing the possibility of desires that are themselves beliefs, then in the crucial cases the claim simply isn’t true. Some beliefs are not compatible with extra desires.

All the same, it may be felt that neither reply tackles the intuition which — perhaps mistakenly — is taken to be validated by the Argument from Dislike. Like the Argument from Doubt, this argument may be a blind alley. But why should it seem appealing in the first place? What is special about desires, in virtue of which they are at least capable of pre-empting other motivations?

In section 3 the analogous question about demonstratives led us to the Targeting Argument — to the idea that in virtue of their perceptual basis, demonstratives play a crucial role in cognitive control of perceptually-guided action. In the present case I think the answer also turns on a special causal role. Desires are being held to fill a causal and explanatory lacuna between beliefs on the one hand, and the Will, on the other.

Desires are the catalysts that cause intentions to condense from beliefs. The Argument from Dislike attempts to establish the need for such a catalyst, by claiming that in conceivable circumstances, a given set of beliefs would not give rise to the formation of a certain intention. We know that the argument fails. But the catalytic view of desire has a lingering appeal.

At this point in the demonstrative and indexical cases, we looked for actual and hypothetical examples to show that the action-guiding causal role of demonstrative modes of presentation could be performed by non-demonstratives. We tried to show that that role is not a kind of cognitive constant, which would have to be tackled in the same way in any conceivable cognitive system. On the contrary, its nature depends very much on contingent features of the system’s environment, its perceptual apparatus, and the symbolic resources of its cognitive centre.

My suggestion now — not a very surprising one, I think — is to adopt the same strategy in the present case, against the catalytic model of desire. That model sees a causal gap between beliefs, on the one hand, and the Will, on the other; intentions cannot condense in the absence of desires. My point is that whether this is so depends on the nature first of beliefs themselves, and secondly of the surface on which intentions are going to form — that is, of the Will. There is no reason to regard these factors as cognitive constants. On the contrary, we shall see that it is easy to imagine agents who differ in these respects, from each other, and from ourselves. In consequence, a non-catalytic intention condenser is far from inconceivable. The interesting question is whether there actually are any.

As in the demonstrative case, there are two main possibilities: where there is a causal gap, it can be closed from either side. (Again I shall ignore the obvious hybrid possibilities, and also solutions involving a separate bridging component.) The first option is to expand an agent’s belief system, so as to admit beliefs that would give rise to intentions directly (or at least as directly as belief-desire combinations are ordinarily held to do so). We want a belief which would in itself normally comprise a reason for performing an action — a belief whose capacity to rationalize an action does not depend on the co-occurrence of certain desires.

I think we can get a good idea of what such a belief would be like, by reflecting on familiar expressions of moral or rational necessity. For example: I must go down to the sea again; or I am bound to tell the truth.
Superficially, at any rate, utterances such as these can be regarded as expressions of a belief that a possible action (going down to the sea, or telling the truth) has a certain property (being what I must do, or am bound to do). Moreover, these beliefs (if they really are beliefs) seem to provide reasons for actions that are insensitive to an agent's desires about actions with the corresponding properties. If I believe that I must go down to the sea again, then it seems I have a reason for going down to the sea even if I do not particularly want to do what I must do. (A man must do what a man must do, whether he wants to or not.) Similarly, if I think I am bound to tell the truth, it seems irrelevant whether I want to do so.

These points are (notoriously) disputable. There are two lines of attack. One urges that believing that I must do something, or am bound to do it, only provides me with a reason for doing it because, like everybody else, I have an implicit desire to do what I must. The other takes the opposite tack, and argues that because utterances such as these express reasons for actions, they do not express beliefs (and hence have to be construed some other way). However, both lines of argument seem to depend on the very claim at issue: that beliefs alone cannot motivate actions.

It is indeed contentious whether these expressions of felt necessity do, as their form suggests, express genuine beliefs. To tackle the question, however, we need a criterion with which to distinguish beliefs from desires. Hume's principle offers such a criterion, saying that desires are distinguished in virtue of their active or motivating role in the production of action. But we now find that the principle itself depends on the criterion it claims to supply.¹⁶

Thus it seems that we have no non-circular reason to deny the possibility of an intrinsically rationalizing belief: a belief about a possible action whose capacity to rationalize that action is independent of an agent's desires. And if we have no reason actually to affirm the possibility, this is because the distinction between belief and desire is itself so obscure (one reason, I think, why the argument benefits from comparison to the relatively clear demonstrative and indexical cases).

I said that as in the demonstrative case, there are two main ways to bridge the gap that Hume's principle claims to find between beliefs and the Will. The second is to imagine the Will itself equipped to handle a somewhat broader range of inputs (just as we had to think of an agent's perceptual system as producing a broader range of outputs). Why shouldn't the Will, or the unit in which an agent forms intentions to act, simply be built to respond to certain beliefs about the properties of actions and states of affairs? A familiar example is the classical economic rationalist, whose actions are based entirely on the profit motive; she forms the intention to do whatever she believes will maximize her net monetary return. It might be objected that the rationalist wants to make a profit. But why should she? If she has desires at all, why shouldn't she be built so that desiring something and thinking it profitable are one and the same?

A similar case is the perfect (reflective) Bayesian, who chooses the action he takes to maximize his expected utility. Whether he has desires concerning possible outcomes or beliefs concerning their utilities, his final intention to act must be based simply on the belief that that action maximizes expected utility. For if we have to add that he wants to maximize expected utility, then in order to do so he has to make a further calculation, taking that want into account. To insist on a catalytic desire would be to prevent him from acting at all.

These cases are admittedly unrealistic. But they show, I think, that there is no conceptual obstacle to the idea of an agent who is motivated to act by beliefs that in our own case would need to be supplemented by other attitudes. We are not economic rationalists or perfect reflective Bayesians. But we can imagine circumstances in which our deliberations and choices would parallel those of such agents. In the rationalist case, for example, we can see that this would happen if we acquired an overriding desire for monetary gain. Modulo that desire, we remain deliberative agents; and hence it is a short step to a conception of an agent in whom the function of that desire is factored out—in whom the Will, the intention-forming psychological component, is simply structured to respond to beliefs of the appropriate sort.

Thus it seems to me that the case for Hume's principle is indeed vulnerable to parallel objections to those we have raised against the arguments of Peacocke and Perry. It derives some unwarranted plausibility from an analogue of the Argument from Doubt. And more importantly, it turns out to depend on unjustified assumptions about the causal connections between the mental 'organs' of a cognitive agent.

The major difference, I think, concerns the bearing of these conclusions on the respective applications of the two sorts of argument. In the indexical case, the conclusion seemed of largely theoretical interest.
There is no doubt that in practice we do depend heavily on demonstrative and indexical modes of presentation of features of our environment. The mistake is to try to generalize on this, extrapolating to arbitrary cognitive systems what are really peculiarities of our own situation.

In the desire case, however, the conclusion lies closer to home. In this case the practical question not only remains unsettled, in the sense that is not clear to what extent our actions are in fact explained by attitudes that are not beliefs; but also the question itself is under threat. Hume's principle is not only taken to show that action required desires, as well as beliefs; it has also provided a reason for distinguishing desires from beliefs. Without it, we lack a foundation for that distinction. As cognitive scientists, this means that we lack a justification for the traditional two-box taxonomy of cognitive states — a reason not to opt for a single species of propositional attitude, differentiated entirely by content.

While as non-cognitivists, it means that having sought to ground our semantic distinction psychologically, we may be forced to retract our steps. Like Hume himself, we may feel bound to say that desires are distinguished from beliefs in virtue of the fact that it is "impossible (.) they can be pronounced either true or false". (Of course, the problem then will be to show that moral commitments are not true or false.)

In either case, the prospect may be so uninviting, or at least so unconventional, that we choose to defend Hume's conclusion. The point of this excursion has been to show that its present defences are rather inadequate. Like many a tradition, it survives at present not so much on established merits as simply in virtue of not being challenged. But nothing challenges tradition like the discovery that others might do things differently. Hence this brief excursion to the possible worlds. In order to counter the tendency to unwarranted generalization from our own case, I have tried to encourage a flexible and speculative approach to the cognitive functions that have seemed to necessitate certain sorts of mental states. What travel broadens, in this case, is one's view of the possible structure of the mind.

NOTES
1 I am grateful for comments from Hugh Mellor, Philip Pettit, Phillip Staines and participants in a seminar at ANU in 1986; also for financial support from a National Research Fellowship.
2 David Hume, Treatise, Book II, Part III, Section III.